SALVAGE Report D2.1 Description of existing and extended smart grid component models for use in the intrusion detection system

Kosek, Anna Magdalena; Heussen, Kai

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

SmartGrids ERA-Net

Project:
Cyber-phySicAl security for Low-VoltAGE grids (SALVAGE)

Project partners:
KTH - Royal Institute of Technology
DTU - Technical University of Denmark
PWR - Wroclaw Institute of Technology

SALVAGE D 2.1

Description of existing and extended smart grid component models for use in the intrusion detection system

Anna Magdalena Kosek (DTU)
Kai Heussen (DTU)

June 2015

Revision history

Issue	Date	Changed page(s)	Cause of Change	Implemented by
0.1	13-03-2015	All	First draft	Anna Magdalena Kosek (DTU)
0.2	23-04-2015	All	Second draft	Anna Magdalena Kosek (DTU) Kai Heussen (DTU)
0.3	03-06-2014	All	Final draft	Anna Magdalena Kosek (DTU)
0.4	26-06-2014	3-10	Final draft with reviewer comments	Kai Heussen (DTU) Matus Korman (KTH)
1.0	30-06-2015	1,4,6,7,8, 11,25,28	Final document	Anna Magdalena Kosek (DTU)

Table of contents

SALVAGE project

The purpose of the SALVAGE project is to develop better support for managing and designing a secure future smart grid. This approach includes cyber security technologies dedicated to power grid operation as well as support for the migration to the future smart grid solutions, including the legacy of ICT that necessarily will be part of it. The objective is further to develop cyber security technology and methodology optimized with the particular needs and context of the power industry, something that is to a large extent lacking in general cyber security best practices and technologies today. In particular the focus of the project will be on smart grid with many small distributed energy resources, in particular LV substation automation systems and LV distribution system.

1 Introduction

This report explores models already available at DTU and investigates modeling techniques usable for model-based anomaly detection of a single power system component connected to the distribution grid (consumer or producer). Section 2 identifies the model requirements for intrusion detection, including the introduction to the SALVAGE framework and the working scenario. The intrusion detection concept from computer science and anomaly detection applied to physical behavior of DERs is introduced in Section 3. Section 4 discusses DER modeling for the purpose of the model-based intrusion detection and introduces models developed by DTU Electrical Engineering, Energy System Operation and Management Group. Section 5 investigates different data-driven PV models usable for model-based intrusion detection and discusses data cleaning, modeling techniques, required model inputs, finally the available PV models are evaluated and compared. Section 6 introduces the proposed method of anomaly based intrusion detection with use of physical PV model including data preparation and cleaning, power production prediction, residual analysis and anomaly evaluation. Section 7 concludes and outlines future SALVAGE research in the area of the cyber-physical model-based intrusion detection.

2 Model Requirements for Intrusion Detection System

Intrusion detection systems (IDS) discern observable events reflected in available data and metrics into information about possible cyber-attacks. In physical infrastructures, computational models of the system's behavior are often used for decision and control purposes, and are thus often readily available. The idea of model-based cyber-physical intrusion detection aims to integrate such behavior models into a framework for detection of external interference (with assumed malicious intent).

In this section we identify what type of information and knowledge can be inferred from behavioral representations of a system and how such models should be integrated support intrusion detection.

2.1 SALVAGE IDS framework

The SALVAGE project investigates cyber-physical vulnerabilities in the distribution network: cyber vulnerability assessment, physical intrusion detection, power system vulnerability assessment, power system impact of cyber-attacks. The overall requirements for the IDS framework in the context of the SALVAGE project are motivated by the following observations:

- Cyber-attacks aimed at disrupting system operation may manifest themselves to the observer in much the same way as system faults, such as normal component failures, : a successful cyber-attack will cause the system behavior to deviate from a desired behavior.
- Cyber-attacks may be designed to cause system failure because that is the desired destructive behavior or to simply emulate behaviors, for cloaking or detection avoidance; both strategies, for example, have been employed in case of Stuxnet [4].
- The ability to correctly distinguish between failure and attack post-mortem is just as important as the ability to prevent attacks in the first place; a recognized and analyzed attack pattern

enables prevention of attack repetition and elimination of attack vectors.

- The engineering effort available for develop accurate simulation models for an IDS facilitating cyber-attack prevention at distribution level is large. Practical IDS development and operation should therefore be largely independent of accurate physical model of the investigated physical system.

The first two observations motivate the use of behavioral models representing the expected cyber-physical system behavior for reference and on-line detection; the latter observations however also encourage an engineering strategy for IDS development that integrates well with other engineering and business work flows and focuses on a close interaction with available data sources (a data-driven engineering approach).

A pure data-driven modeling itself cannot be sufficient to discriminate between the various types of attack vectors. In addition to behavioral representations, also a qualitative framework is required to hypothesize and evaluate alternative attack scenarios.

In the SALVAGE framework, we therefore propose a two-layered approach as illustrated in Figure 1. The analysis of measurements and integration of cyber-physical behavioral models aims toward *anomaly detection*; the interpretation of observed anomalies is then passed to another analysis layer aimed at *intrusion detection*, which integrates anomaly information with knowledge about expected system behaviors (for example control signals, current system configurations, operating modes and procedures).

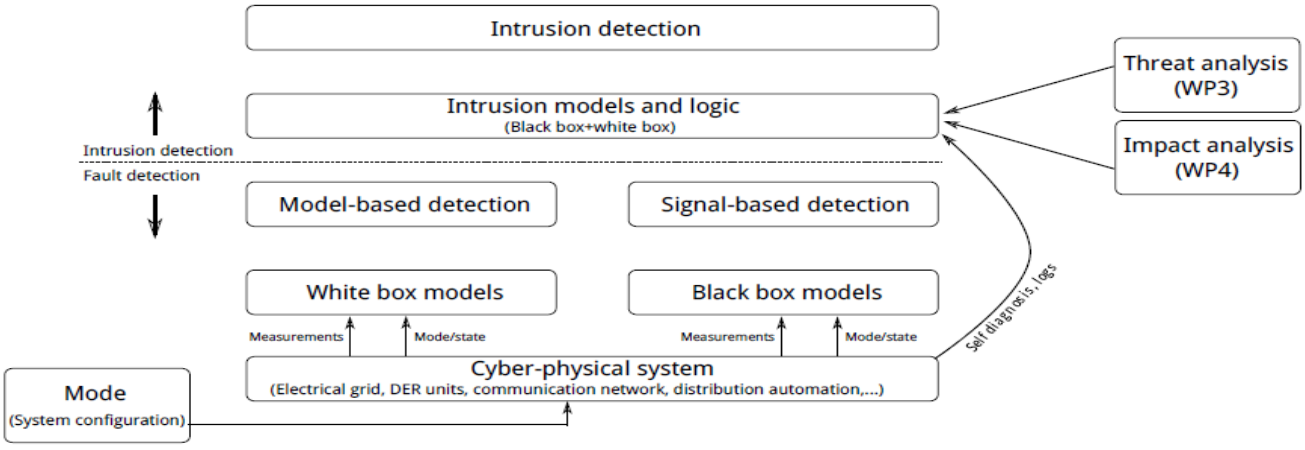

Figure 1 Outline of SALVAGE Intrusion detection framework.

Behavioral models represent a previously observed and explained behavior; in case of cyber-physical models, this behavior is typically the dynamic behavior of a physical system which is under some control influence. This behavior can be employed serve as a normative reference (‘normal behavior’), and in this way they may be employed as a normative reference for cross-validating measurements to detect deviations from the modeled outputs as behavior anomalies. Whereas such models represent physical and controlled system behavior well, they contain no information about possible intrusion pathways, thus they may only serve to detect behavior anomalies, but do not identify the ‘intent’ in the behavior.

As part of the SALVAGE efforts, this report present the anomaly detection part of Work Package 2,

Model-based intrusion detection, investigates intrusion detection methods based on the attack scenarios outlined in WP1. The research and experimental work includes white and a black box model based on measurements and physical system state, and investigates model-based and signal-based intrusion detection. The presented first part of WP2 work is an investigation of anomaly detection techniques for distinguishing normal and suspicious operation, as described in this report.

2.2 SALVAGE working scenario

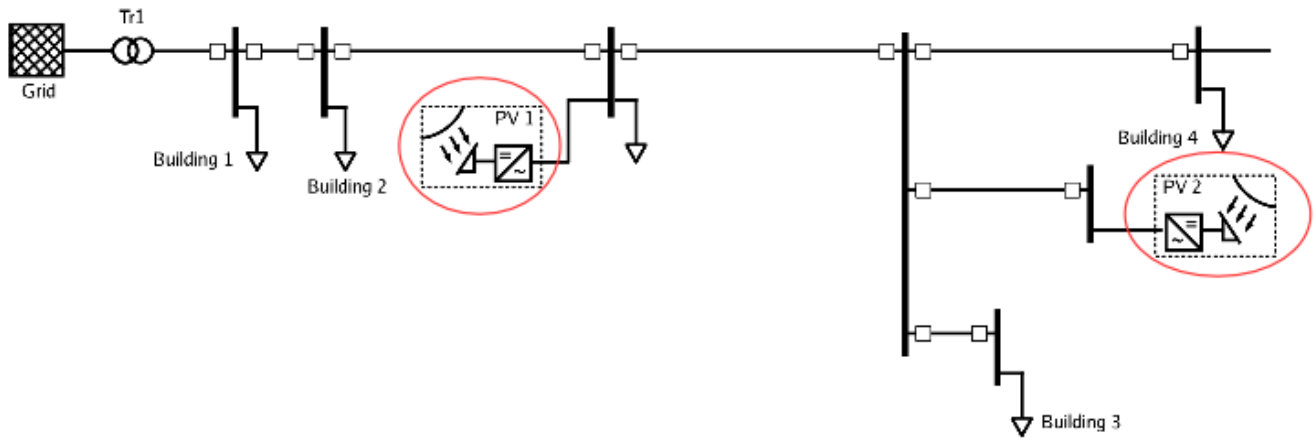

Figure 2: SALVAGE PowerCap scenario as presented in [2]

As a development context for the SALVAGE techniques, a number of scenarios are developed in WP1. We assume the proposed intrusion detection system to be situated within the SALVAGE distribution network scenario outlined in [2]. The scenario considers a low voltage distribution grid, as presented in Figure 2, with buildings and PVs connected to the transformer (Tr1). In this report we are focusing on the independent components and representing them with models, therefore we are considering cyber-security aspects of buildings and PVs.

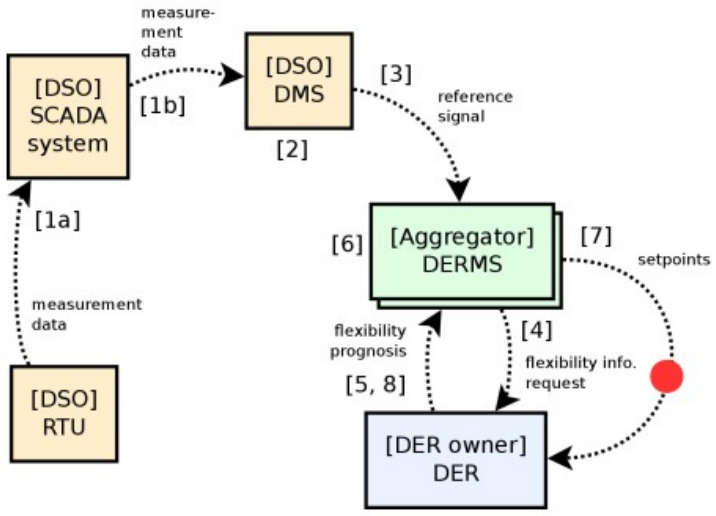

Figure 3.The interactions between DER and Aggregator in the considered scenario [2].

The DER components considered in this report can be externally controlled, for example from the Aggregator. As shown in figure 3, the interaction (noted with red dot) refers to sending set-points from

the Aggregator to the DER unit. The security of this interaction considers an intruder overwriting the setting originating at the Aggregator or setting an independent set-point to the unit.

3 Intrusion detection

Intrusion detection system (IDS), in computer science, gathers and analyzes the information form the computer network or system activities in order to discover malicious activities or violations of policy. IDSs use one of two detection techniques:

- Statistical anomaly based IDS - where the anomalies are detected by comparing the system or the network behavior with the established baseline behavior. The baseline behavior is referred to a 'normal' behavior, where the intrusion is defined as anomaly or significantly different than the baseline.
- Signature-based IDS - where known intrusions or attacks are recorded or defined with a signature and the network or the system behavior is compared to these signatures in order to find the match to well-known malicious behavior and intrusions.

Anomaly based intrusion detection is explored for malicious control of DER in power system. The complex physical behavior of the unit ('normal' behavior) is obtained from the DER model and compared to the observed behavior in order to detect anomalies. This way the IDS can divide the observed behavior of the unit into 'normal' and 'suspicious'. In order to detect the reason of the suspicious behavior, signature-based IDS can be used to detect for example device failures.

In the WP2 of the SALVAGE project we first consider anomaly detection with use of DER models and continue to signature-based intrusion detection to explore the reason for anomaly, focusing on distinguishing between verified and malicious control. This report describes the model-based intrusion detection technique of a single DER connected to the distribution grid.

3.1 Anomaly detection applied to physical behavior models

In order to verify if the unit is being controlled by an unauthorized entity, its behavior can be observed and classified into four categories:

- normal operation- when the unit behaves as expected and it is not controlled by any external set-point,
- faulty operation – when the unit's operation is distorted by a fault on the unit or in its electrical network environment,
- verified control – when unit behaves as expected under a verified control scheme, or according to issued set-points,
- malicious control – the unit is operated with unverified control scheme or unauthorized set-points.

The prosed categorization of the unit behavior can be included in process of detecting an intrusion affecting the operation of a DER. In order to discover anomalous behavior of the system, the ‘normal’ behavior needs to be described. When detecting anomalies in behavior of a DER, for example their power production or consumption, its physical model can be used to define their ‘normal’ behavior.

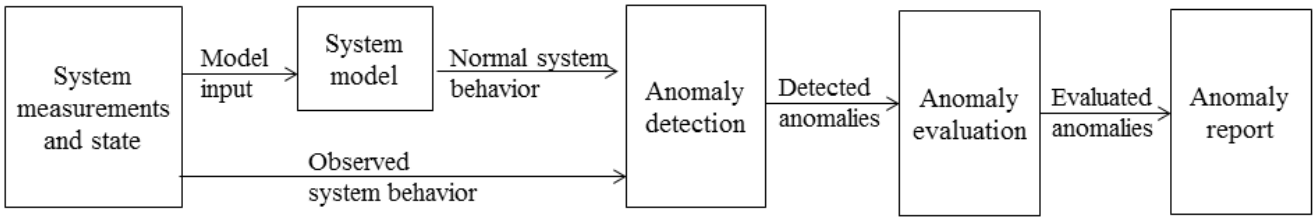

Figure 4. Anomaly detection based on physical behavior models.

The proposed IDS uses anomaly detection method where the model output and the measurement data is compared and their difference is analyzed, as presented in Figure 4. In order to perform the anomaly detection the normal behavior of the DER need to be defined as a DER model, this model takes measurements and DER state and outputs data associated with the normal behavior (for example an active power value of a power producer). The difference between the normal and observed behavior is weighted in order to detect anomalies. Next anomalies are evaluated and the IDS outputs a report about discovered anomalies.

4 DER Modeling

This report investigates the anomaly-based intrusion detection with use of model applied to individual power system components: DERs. DER models need to be for the proposed anomaly-based intrusion detection. This chapter presents selected DER models developed by DTU Electrical Engineering at Energy Systems Operation and Management Group. The existing models are described, categorized and evaluated by Evaluate types by, complexity of modeling tasks, behavior modes, available redundant input, relevance and impact for grid, and use for the selected SALVAGE scenario. Data driven modeling is presented in this section and its advantages and disadvantages are highlighted.

4.1 Discussion of Existing Models

Appendix 1 presents a list of selected power system components and grid models existing in SYSLAB laboratory. The model list and characteristics are included in Appendix 1. Types of modeled units presented in Table 5 in Appendix 1 are as follows:

- PV - photovoltaic array
- Battery
- Grid - subset of a distribution grid
- House - heating system or the building thermal model
- Fridge
- Washing machine
- Microwave

The presented models are evaluated in several aspects, relevant to the SALVAGE project:

- **Complexity of modeling tasks** – how complex is the component model, what is the effort to create an individual model and a set of models representing the same DER type
- **Behavior modes –** ability of the model to recognize different types of behavior, for example normal behavior of an occupied or unoccupied house.

- **Available redundant input** – does the model uses redundant data in order to eliminate the anomalies created by faulty sensor readings
- **Relevance and impact on the distribution grid** – how large is the impact of the malicious control or behavior of the modeled unit
- **Use for scenario** – does the modeled unit appear in one of the SALVAGE scenarios.

Table 1 presents the evaluation of models presented in Appendix 1 and presents the basis for the final model selection.

Table 1. Evaluation of existing DER models.

DER type Evaluation	PV	Battery	Grid	House	Fridge	Washing machine	Microwave
Complexity of modeling tasks	Complex	Complex	Complex	Complex	Simple	Simple	Simple
Behavior modes	No	No	No	No	No	No	No
Available redundant input	No	No	No	No	No	No	No
Relevance and impact on the distribution grid	Relevant	Relevant	Very relevant	Relevant	Low impact	Low impact	Low impact
Use for scenario	Scenario 1	Not considered	Scenario 1	Scenario 1	Not considered	Not considered	Not considered

Based on the evaluation presented in Table 1, models of PV, battery, grid and house are the most relevant models, considering the impact of the malicious operation on the distribution grid. Only PV, grid and house models are relevant for the SALVAGE Scenario 1. This report considers DERs, therefore the focus can be put on a PV or a house model. The PV model was chosen for the first testing of intrusion detection system, as its operation only depends on the environmental conditions and external control signals. The house is a much more complicated case as it also includes the human interaction that is not easily modeled and predicted.

The existing physical models of the laboratory PV actually refer just to one PV in SYSLAB laboratory (see [5] and section 5), out of three. This PV was carefully measured and identified in order to create very accurate model, used for PV use optimization. The existing PV model cannot be reused for other PVs in the SYSLAB laboratory. A large modeling effort is required to obtain models for all available PVs in SYSLAB. Since the IDS should be applied to a large number of PVs, accurate models might not exist for all intended PVs. In this project we investigate the use of data driven methods and machine learning to determine the PV model based on the observed historical data.

4.2 Data driven modeling

Data driven modeling refers to modeling efforts based on historical observations of the considered system. This type of modeling can be used for black-box modeling, where only input and output of the system is observable, and the internal workings are unknown. Data-driven modeling is common in data

mining and machine learning. Machine learning is a method of programming computers to act in a way that have not been explicitly programmed. Data mining is a computational process aiming at discovering patterns in data.

The advantage of using data driven modeling is its flexibility: this computational method can identify the observed unit without a prior knowledge of its inner workings. Models can be created from historical observations of any system. Even unknown patters and complex phenomena can be discovered in the data and expressed in the model. Model can be calculate automatically and updated if the system behavior changes. Disadvantages include large computation effort to generate the model, the model quality depends on the data quality and generality of the model depends on the statistical properties of the data and the data size.

5 PV modeling

This section presents PV modeling efforts with data driven approaches. Section 5.1 presents the PV modeling strategy used for all presented models, including data cleaning, aggregation, selection and pre-processing. Next four regression models (linear or polynomial) based on meteorological data are presented in section 5.2. Sections 5.3 and present two artificial neural network models for the same PV in SYSLAB: meteorological model and neighborhood model. The presented models are validated and compared in Section 5.5.

Figure 5. SYSLAB laboratory at DTU Risø Campus.

SYSLAB [5] is a research facility for intelligent, active and distributed power systems at Technical University in Denmark, Risø Campus, as presented in Figure 5. SYSLAB enables research and testing of control concepts and strategies for power systems with distributed control and integrating a number of decentralized production and consumption components including wind turbines and PV plant in a systems context. SYSLAB consists of three interconnected sites. It includes two wind turbines (11kW and 55kW), a PV plants (7 kW, 10kW, 10kW), a diesel generation set (48 kW / 60 kVA), the PowerFlexHouse Complex (3 houses of 20kW), a 15 kW / 120 kWh vanadium redox flow battery, and a number of loads (75 kW, 3 x 36 kW). They are all connected in one distributed control and

measurement system that enables very flexible setup with respect to experimental configuration. It allows grid connected operation, island operation, or operation in parallel with wind turbine or PV plant.

In order to create models in this section data form meteorological station, and three PVs: PV117, PV319, PV715, asp resented in section 5.2.

5.1 PV modeling strategy

The raw data from the laboratory need to be prepared before it can be used for modeling. The stages of the chosen data driven model development are presented in Figure 6.

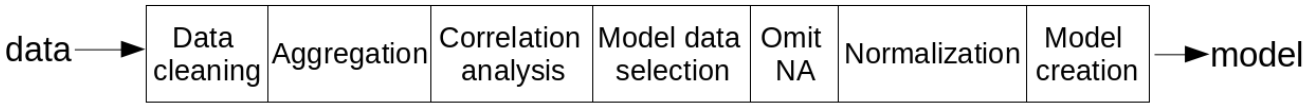

Figure 6. Developed data driven modeling strategy for ANN models.

The data cleaning, preparation and pre-processing stages are as follows:

1. **Data cleaning:** There can be many errors appearing in the raw sensor data. The process of cleaning the data starts form filling missing time series with NA values, so all time-series for a day consist of exactly 86400 data points. The next step is discovering if the time-series have unexpected values. The threshold between realistic and unrealistic values can be determined manually. For example temperature of 50°C is unrealistic for Danish outdoor temperature and can be discarded. Unrealistic values are replaced with NA. If the number of unrealistic values is large the sensor might have been broken. In this case we use signal processing to remove the noise form the data and replace it with NA.
2. **Aggregation:** Once the data is clean and uniform size it can be aggregated. One second values are aggregated to 1 minute values, reducing the size of time-series to 1440 samples. This process is performed in order to save processing time for model creation, while not losing much of information. The aggregation is done omitting NA and with mean value for samples.
3. **Correlation analysis:** The correlation of model output and input is calculated. If the correlation is very small (*correlation<0.2*) the entire day is removed from the data. This step eliminates data with standard deviation equal to zero (which in unrealistic for sensor data) which are periods with a long sensor failures. The statistical significance is not considered here.
4. **Model data selection:** The days secreted by the previous step have been used for the model creation. The timestamps are removed and data samples are grouped in vectors, where sample *I* is as follows: $sample^{(i)} = \left(input_1^{(i)}, input_2^{(i)}, ..., input_n^{(i)}, output^{(i)}\right)$.
5. **Omit NA:** All samples where at least one value in the vector $sample^{(i)}$ is equal to NA are omitted, as used modeling method, ANN, does not accept vectors with NA values. The set of observations $S = sample^{(1)}, sample^{(2)}, ..., sample^{(k)}$ is divided into three groups of random samples: training set S_T, cross-validation set S_{CV}, and validation set S_V, so that $S = S_T \cup S_{CV} \cup S_V$.
6. **Normalization:** Vector normalization is usually performed before ANN model fitting. Samples from S_T have been normalized.
7. **Model creation:** The model was created using one of two ANN libraries for R (`nnet` and

`neuralnet`) with parameters adjusted to fit the training data.

The following sections present simple linear and polynomial regression meteorological models (section 5.2), ANN meteorological model is described in section 5.3, and section 5.4 presents ANN neighborhood model, all predicting active power output of a PV. All models are trained in the same set of 1 second time-series data, power production of PVs in SYSLAB laboratory and meteorological data from October 2014. The data used to train models from the following sections is 1 minute time-series consisting of 44640 rows, randomly divided into 3 sets: S_T, S_{CV}, S_V of size 14841, 14901, and 14898 accordingly.

5.2 Regression meteorological model

Several simple linear regression models were created to predict the active power production of the considered PV. Presented meteorological models take solar irradiation, wind speed, wind direction and ambient temperature and output the expected power production in kW, as presented in Figure 7.

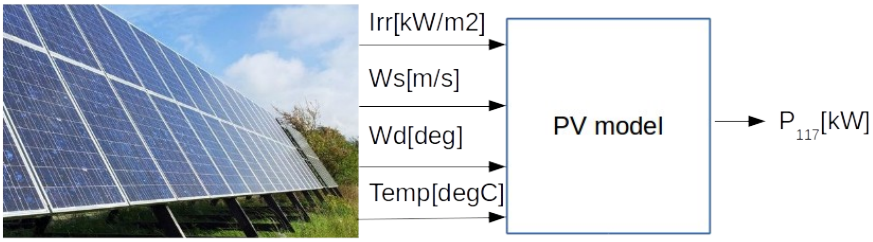

Figure 7. Simple regression PV model graphical representation.

Basic R library `lm` was used to create linear and polynomial models from input data. In this section we preset five models and compare them.

Proposed hypothesis h_0, h_1, h_2, h_3 for construction of five models M_0, M_1, M_2, M_3 are as follows:

$$h_0\left(\theta^{(1)}\right) = \theta_0^{(1)} + \theta_1^{(1)} Irr$$

$$h_1\left(\theta^{(1)}\right) = \theta_0^{(1)} + \theta_1^{(1)} Irr + \theta_2^{(1)} Ws + \theta_3^{(1)} Wd + \theta_4^{(1)} Temp$$

$$h_2\left(\theta^{(2)}\right) = \theta_0^{(2)} + \theta_1^{(2)} Irr + \theta_2^{(2)} Ws + \theta_3^{(2)} Wd + \theta_4^{(2)} Temp + \theta_5^{(2)} Irr^2 + \theta_6^{(2)} Ws^2 + \theta_7^{(2)} W^2 + \theta_8^{(2)} Temp^2$$

$$h_3\left(\theta^{(3)}\right) = \theta_0^{(3)} + \theta_1^{(3)} Irr + \theta_2^{(3)} Ws + \theta_3^{(3)} Wd + \theta_4^{(3)} Temp + \theta_5^{(3)} Irr^2 + \theta_6^{(3)} Ws^2 + \theta_7^{(3)} Wd^2 + \theta_8^{(3)} Temp^2 + \theta_9^{(3)} Irr^3$$

Parameters for each hypothesis $\theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \theta^{(3)}$ were calculated with use of training set S_T of 14841 samples (out of 44640 samples form the *S* set) with use of `lm` function in R.

5.2.1 Regression model-0

Takes irradiation to calculate the power consumption with formula as follows:

$h_0\left(\theta^{(1)}\right) = \theta_0^{(1)} + \theta_1^{(1)} Irr$, where $\theta^{(0)} = (\ 0.1094376 \quad 2.2604230\ )$

Figure 8 shows the linear, single variable model for the PV production, data points are model input data form set S_T and the power calculated form this set. Figure 9 presents the model prediction (red)

compared to the real production data (black) form set S_{CV} . It is visible that the model is not very accurate in minimums and maximums of the production, therefore a simple model cannot be used for the intrusion detection purposes.

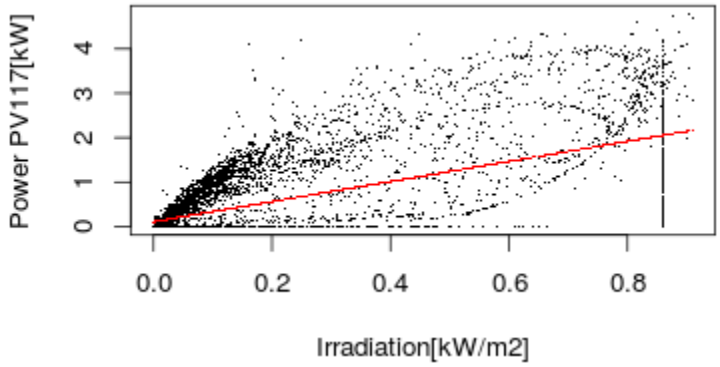

Figure 8. Model-0 training data and model output.

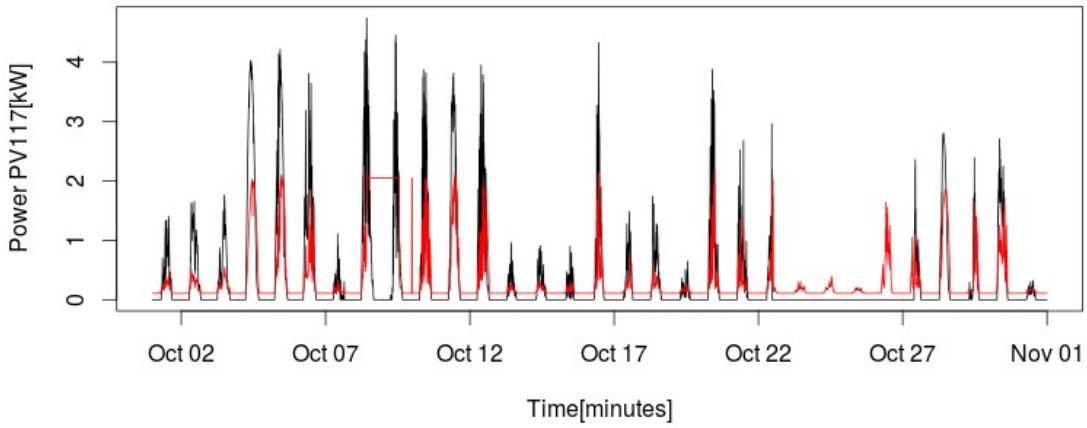

Figure 9. Model-0 output compared to the measured PV117 production.

5.2.2 Regression model-1

Takes irradiation, wind speed, wind direction, ambient temperature to calculate the power consumption with formula as follows:

$h_1\left(\theta^{(1)}\right) = \theta_0^{(1)} + \theta_1^{(1)} Irr + \theta_2^{(1)} Ws + \theta_3^{(1)} Wd + \theta_4^{(1)} Temp$, where $\theta^{(1)}$ =(*0.0152916763304, 2.6721188099024, 0.0689680031581,-0.0001766575963, 0.0014208500330)*

Figure 11 shows the linear model for the PV production, data points are model input data form set S_T and the power calculated form this set.

Figure 11 presents the model prediction (red) compared to the real production data (black) form set S_{CV} . In comparison to the output for Model-0 (Figure 9), model-1 predicts better in the minimums, and performs slight better in maximums.

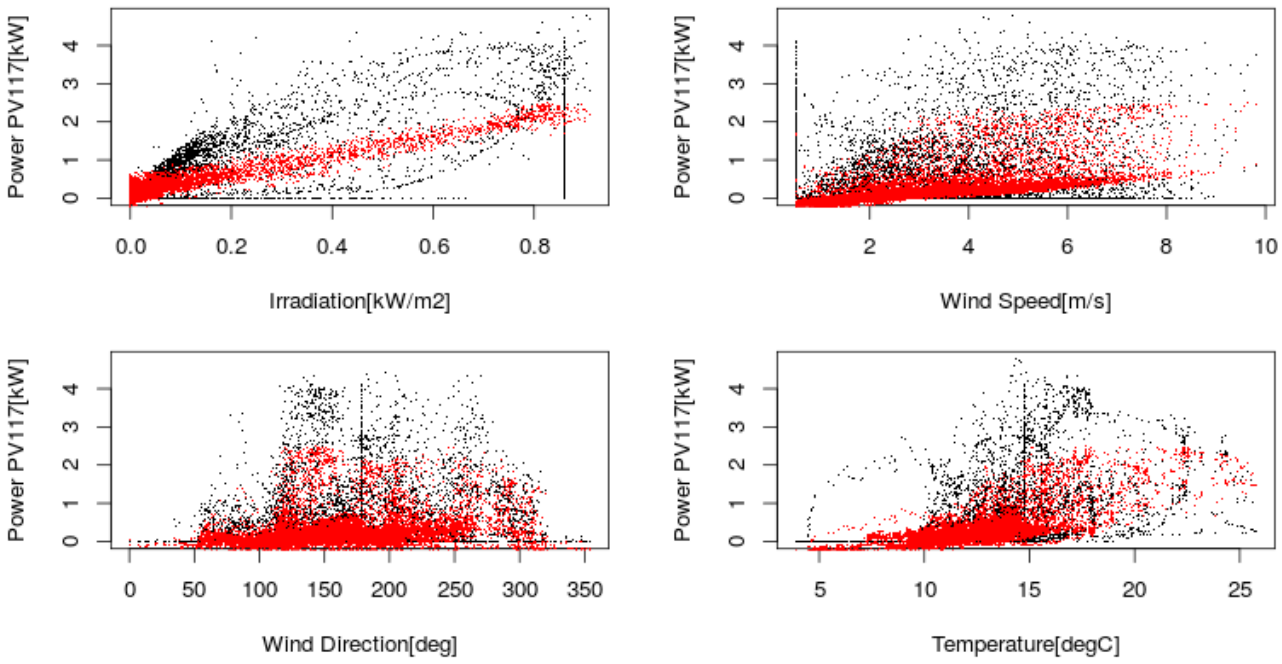

Figure 11. Model-1 training input and output data.

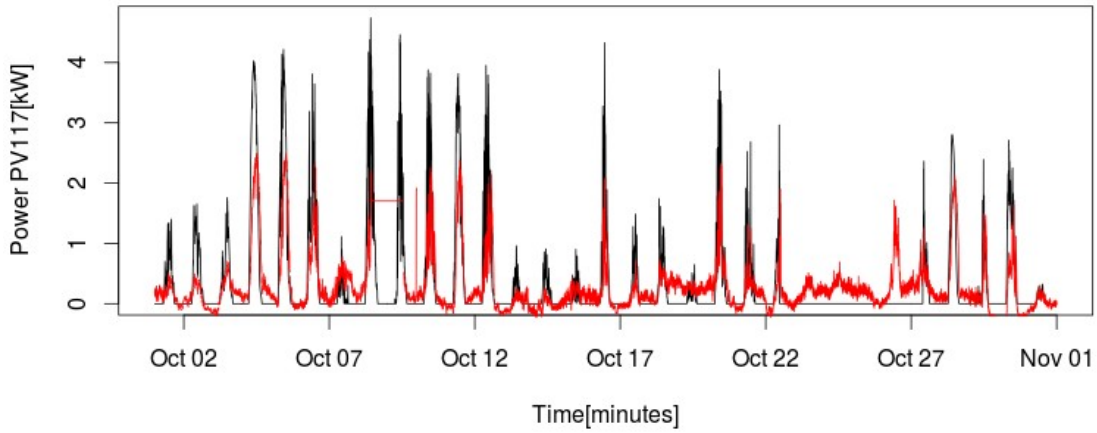

Figure 10. Model-1 PV power prediction results (red) obtained with the cross-validation data set.

5.2.3 Regression model-2

This linear model takes irradiation, wind speed, wind direction, ambient temperature to calculate the power consumption with formula as follows:

$h_2\left(\theta^{(2)}\right) = \theta_0^{(2)} + \theta_1^{(2)} Irr + \theta_2^{(2)} Ws + \theta_3^{(2)} Wd + \theta_4^{(2)} Temp + \theta_5^{(2)} Irr^2 + \theta_6^{(2)} Ws^2 + \theta_7^{(2)} W^2 + \theta_8^{(2)} Temp^2$, where

$\theta^{(2)}$ = *(-0.990668284934828, 6.070479844300378, 0.110911311188601, -0.003277889451628, 0.166264220499390, -4.012042544422539, -0.006648793318575 , 0.000006192536912, -0.006105482360976)*

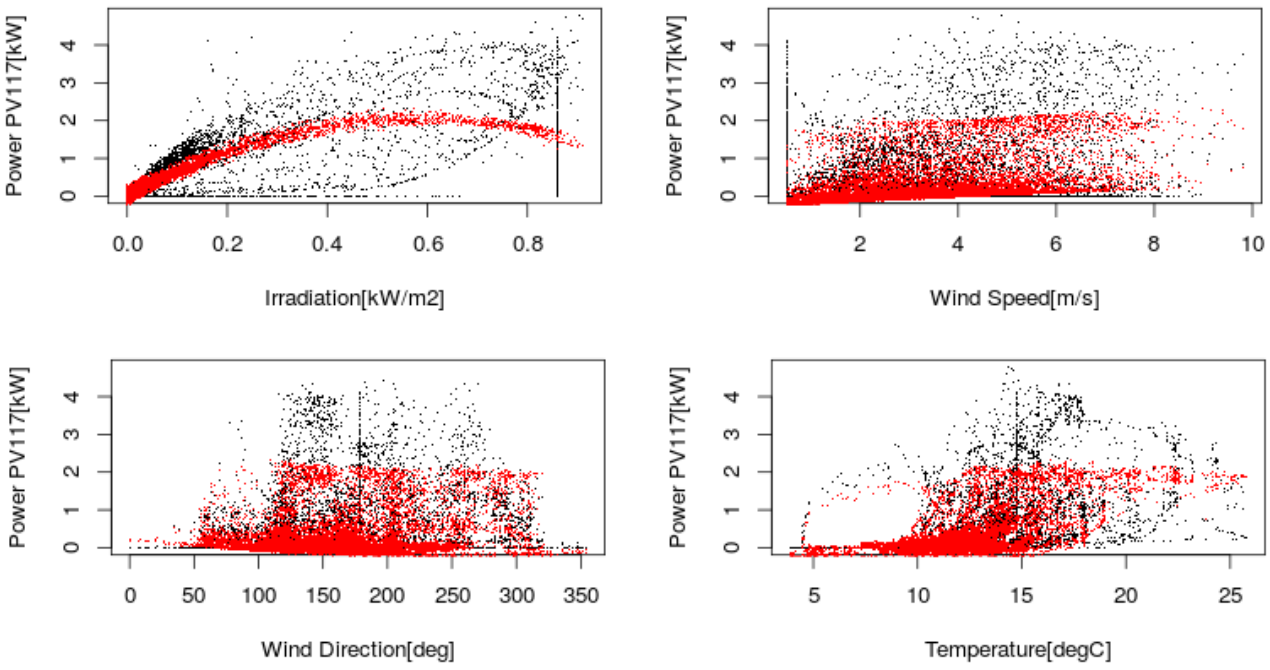

Figure 12. Model-2 graphical representation on the training data.

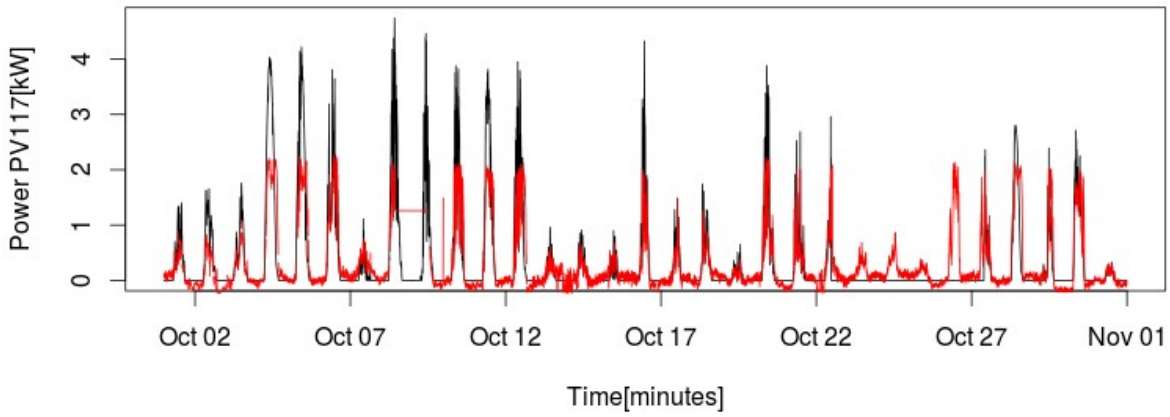

Figure 13. Model-2 PV power prediction results (red) obtained with the cross-validation data set.

Relations between polynomial model presented in this section and the training input data S_T is presented in the Figure 12. The parabolic shape of the power prediction mapped to the irradiation, does not reflect the strong correlation between the data, and therefore suggests that the power production maximums will be decreased in the model.

Wind direction and temperature are better mapped by the model to the power production, in comparison to model-0 (Figure 8) and model-1 (Figure 11), improving the power prediction only slightly, as shown in .

5.2.4 Regression model-3

This model takes irradiation, wind speed, wind direction, ambient temperature to calculate the power consumption with formula as follows:

$h_3\left(\theta^{(3)}\right) = \theta_0^{(3)} + \theta_1^{(3)} Irr + \theta_2^{(3)} Ws + \theta_3^{(3)} Wd + \theta_4^{(3)} Temp + \theta_5^{(3)} Irr^2 + \theta_6^{(3)} Ws^2 + \theta_7^{(3)} Wd^2 + \theta_8^{(3)} Temp^2 + \theta_9^{(3)} Irr^3 + \theta_{10}^{(3)} Ws^3 + \theta_{11}^{(3)} Wd^3 + \theta_{12}^{(3)} Temp^3$,where $\theta^{(3)}$ = (8.815803e-01, 5.244864e+00, 6.418637e-01, -1.302377e-03, -3.323558e-01, 6.125233e-02, -1.589988e-01, -9.234099e-06, 2.597408e-02, -4.634485e+00, 1.249784e-02, 3.622099e-08, -6.469007e-04).

The model input (black) and output for irradiation, wind speed, wind direction and ambient temperature is presented in Figure 14. In comparison with model-2 (Figure 12), model-3 behaves similarly; also inheriting the problem with predicting maximums of the power production, see Figure 15.

The additional variables do not improve the model-1 in comparison to model-2 (Figure 12). A nonlinear model should be explored to improve the prediction.

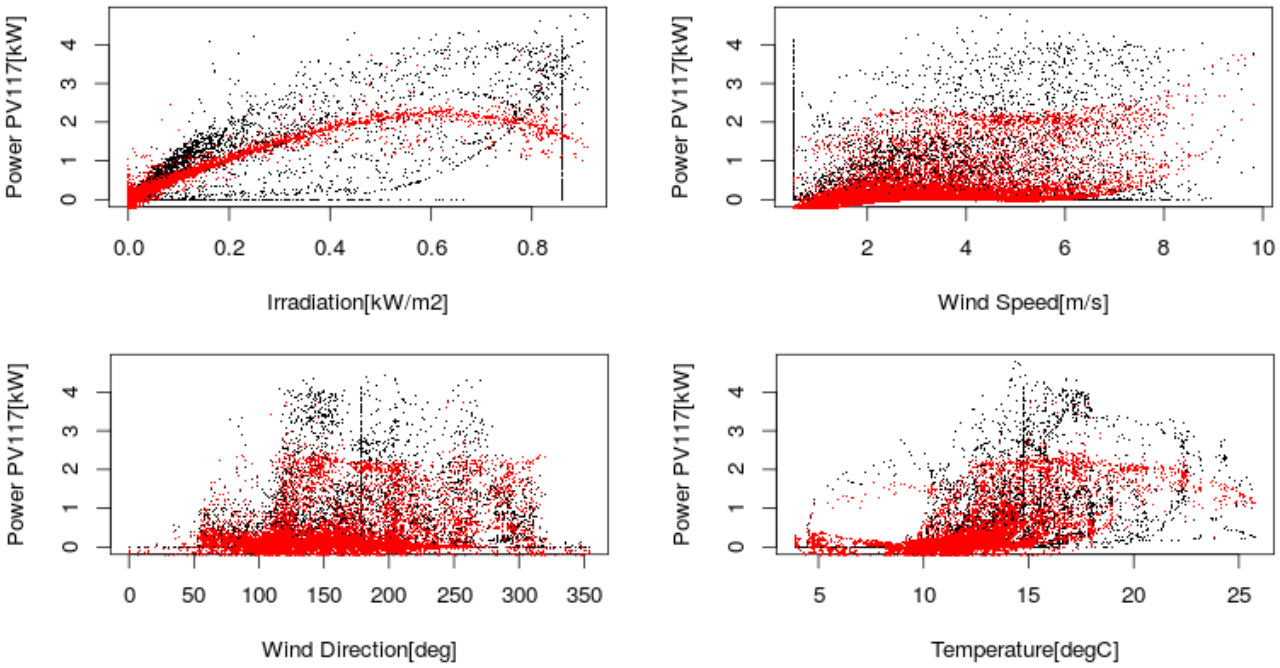

Figure 14. Model-3 graphical representation on the training data.

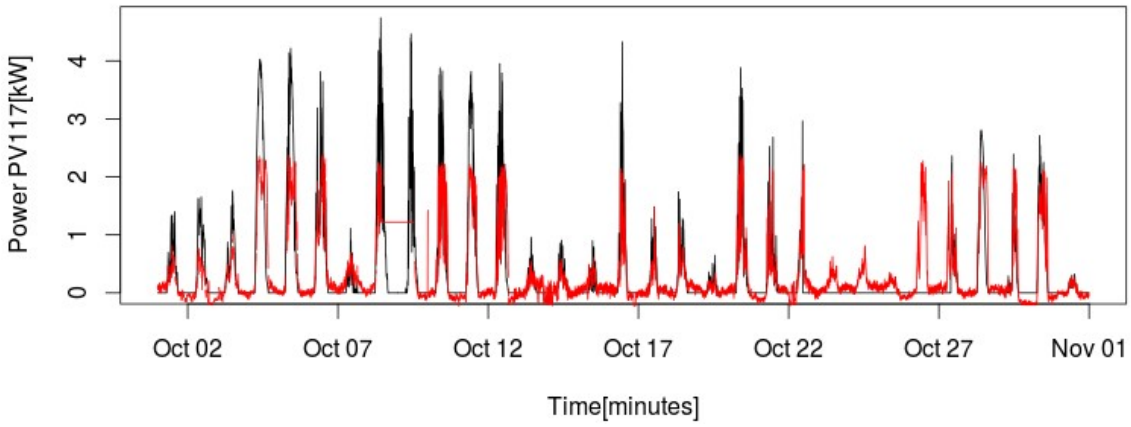

Figure 15. Model-3 PV power prediction results (red) obtained with the cross-validation data set.

-----Pages 16-25 not available-----

Pages 17-26 contain results that are part of a scientific publication that is currently under review, these pages will be available after March 2016. In case of questions, please contact the corresponding author Anna Magdalena Kosek at amko@elektro.dtu.dk.

7 Conclusions and future work

This report presents model requirements for intrusion detection (section 2) including the SALVAGE IDS framework (section 2.1) and a short description the considered SALVAGE scenario (section 2.2). Anomaly based intrusion detection method applied to physical behavior of a DER is presented in section 3. The discussion about the existing PV operation models and approached to PV modeling are presented in sections 4and 5. Data driven models using linear and polynomial regression and artificial neural networks with different inputs are presented in sections 5.2-5.4, the presented models are compared and evaluated in section 5.5. The proposed model-based intrusion detection system is presented in section 6, and applied to anomaly detection and evaluation of a PV operation with use of presented ANN models.

The future work includes further investigation into formal definition of normal behavior of a PV array, residual analysis and anomaly evaluation. Next signature-based intrusion detection can be used to distinguish control action performed on a PV with use of data from SYSLAB laboratory.

Appendix 1

Table 5. List of selected DER models.

No	Device type	Description	Input		Output	Analytical/ Computational	Quality			Time resolution	Programming language	Link/publication
			Controllable	Uncontrollable			RMSE	variance	validated			
1	House	Flexhouse1heating single room model, discrete space state	Heat power [W]	Solar rad [W/m2], External temperature [C]	Room temperature [C]	Analytical	na	na	yes		R	
2	House	Flexhouse1heating single room linear model, discrete space state	Heat power [W]	Solar rad [W/m2], External temperature [C]	Room temperature [C]	Analytical	na	na	yes		Matlab, Java	
3	House	Flexhouse1heating single room linear model, discrete space state	Heat power [W]	Solar rad [W/m2], External temperature [C]	Room temperature [C]	Analytical	na	na	yes		R	
4	House	Flexhouse1heating single room non-linear model, discrete space state	Heat power [W]	Solar rad [W/m2], External temperature [C], wind speed [m/s]	Room temperature [C]	Analytical	na	na	yes		R	
5	House	Flexhouse 1 heating multiple (8) room model with 10 heathers	10xHeat power [kW]	Solar rad [W/m2], External temperature [C]	8xRoom temperature [C]	Analytical	na	na	yes	variable	Matlab, Python	
6	House	house heating model, adjustable time step and hose size configuration			P, P1-3, Q, Q1-3, Frequency, Voltage, Current, house state, wind speed, wind direction, solar read	Analytical				I sec	Java	risoe.syslab.control.resemu.model.HouseSimulator in syslab pvsec2012
7	PV	SYSLAB PV 319	Ambient temperature [C], solar rad W/m2] wind speed [m/s]		power [kW], theoretical Q, 5 min prediction of P and Q	Analytical					Java	risoe.syslab.control.experiment.pvsec12.PVPotential in syslab pvsec2012
8	Grid model										PowerFactory/ Matlab	
9	Battery	SYSLAB Vanadium battery	voltage of the battery [Volt] SOC[%]	external temperature [C]	P,Q	Analytical		available, to be computed	yes	1 sec	Matlab/ Simulink	http://www.sciencedirect.com/science/article/pii/S0378775313020570
10	Fridge	single state model	electric power [W]	room temperature [C]	refrigerator temperature [C]	A	available, to be computed	available, to be computed	yes	Variable (1-10 min)	Matlab	
11	House	single room thermal model	cooling power[W]	external temperature[C] solar radiation [W/m2]	internal room temperature[C]	Computational	available, to be computed	available, to be computed	yes	Variable (1-5 min) depend of the input data resolution	Python	
12	PV	SYSLAB PV 319		solar radiation[W/m2] outdoor temperature[C], wind speed [m/s], time [h]	power[W]	Computational	available, to be computed	available, to be computed	yes	Variable (1-5 min) depend of the input data resolution	Matlab	
13	Washing machine	Kitchen 776	start/stop, load[full load, half-load, small load]		power [W]	Computational			not (time series of real consumption)	Variable (min 1 sec)	Simulink	
14	Microwave	Kitchen 776	start/stop, load[full load, half-load, small load]		power [W]	Computational			not (time series of real consumption)	Variable (min 1 sec)	Simulink	
15	PV	PV319	Ambient temperature [C], solar rad W/m2] wind speed [m/s]		power [kW], theoretical Q, 5 min prediction of P and Q	Analytical	available, to be computed	available, to be computed	yes	1 sec	Matlab/ Simulink	ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6715023

References

[1] Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0, on-line: http://cran.r-project.org/web/packages/nnet/index.html

[2] Korman, Matus and Mathias, Ekstedt and Kosek , Anna Magdalena *Deliverable 1.1 Smart grid scenario: Project: Cyber-phySicAl security for Low-VoltAGE grids (SALVAGE)* 2015, on-line: http://orbit.dtu.dk/en/publications/deliverable-11-smart-grid-scenario%2830ca1f66-51a3-4c31-835a-848cdcaab078%29.html

[3] Zambrano-Bigiarini , Mauricio *R library hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series*, Version 0.3-8, 2014-02-04 , n-line: http://cran.r-project.org/web/packages/hydroGOF/index.html

[4] Langner, Ralph. "Stuxnet: Dissecting a cyberwarfare weapon." *Security & Privacy, IEEE* 9.3 (2011): 49-51.

[5] SYSLAB laboratory in PowerLabDK on-line: www.powerlab.dk/facilities/syslab.aspx